correlation

Matt Dube

December 15, 2018

library(readr)  
library(dplyr)  
library(here)  
library(sjPlot)

Load data

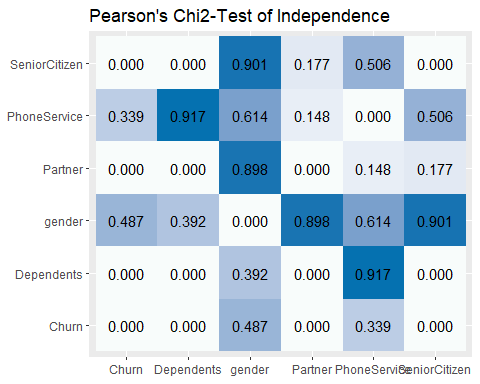
customer <- read\_csv(here("00\_Data/raw", "WA\_Fn-UseC\_-Telco-Customer-Churn.csv"))

chisq.test(customer$Churn, customer$gender)

##   
## Pearson's Chi-squared test with Yates' continuity correction  
##   
## data: customer$Churn and customer$gender  
## X-squared = 0.48408, df = 1, p-value = 0.4866

customer\_twoLevels <- customer %>%  
 mutate(SeniorCitizen = ifelse(SeniorCitizen == 1, "Yes", "No")) %>%  
 select(gender, SeniorCitizen, Partner, Dependents, PhoneService, Churn)

sjp.chi2(as.data.frame(customer\_twoLevels), title = "Pearson's Chi2-Test of Independence")



customer\_clean <- customer %>%  
 filter(!is.na(TotalCharges)) %>%  
 select(-c(customerID, TotalCharges, gender, PhoneService)) %>%  
 mutate(SeniorCitizen = ifelse(SeniorCitizen == 1, "Yes", "No"),  
 Churn = as.factor(Churn)) %>%  
 mutate\_if(is.character, as.factor)